5. MODEL CALIBRATION AND VALIDATION PROCEDURES 5-1

LIST OF FIGURES

Title **Page**

Figure 5-1 Distribution Screen for Point-Source Loading of Toxicant in Water 5-12

Figure 5-2 Uniform Distribution for Henry's Law Constant for Esfenvalerate 5-13

Figure 5-3 Triangular Distribution for Maximum Consumption Rate for Bass.................. 5-13

Figure 5-4 Normal Distribution for Maximum Photosynthetic Rate for Diatoms 5-14

Figure 5-5 Latin Hypercube Sampling of a Cumulative Distribution with a Mean of 25 and Standard Deviation of 8 Divided into 5 Intervals.................................... 5-14

LIST OF TABLES

Title **Page**

Table 5-1 EFDC Model Calibration/Validation Steps 5-4

Table 5-2 Calibration and Validation Periods 5-9

5. MODEL CALIBRATION AND VALIDATION PROCEDURES

5.1 OVERVIEW

All model applications typically include three primary phases or steps: database development, system characterization, and calibration and validation. Section 3 described the general data availability and the conceptual model for the PSA. The data requirements specific to each model were described in Section 4.2, and Section 4.3 presented a discussion of the physical domains that characterize the systems represented by each model. This section provides an overview of the calibration and validation procedures for each model. More detail on the procedures that will be followed during model calibration and validation is provided in the Modeling Study QAPP (Beach et al., 2000).

Model calibration and validation are a necessary and critical step in any model application. Calibration is an iterative procedure of parameter evaluation and refinement, comparing simulated and observed values of interest. Model validation is in reality an extension of the calibration process. The purpose of validation is to ensure that the calibrated model properly represents all the variables and conditions that can affect model results. Model credibility is based on the ability of a single set of parameters to represent the entire range of observed data.

While there are several approaches to calibrating/validating a model, perhaps the most effective procedure is to use only a portion of the available record of observed values for calibration; once the final parameter values are developed through calibration, simulation is performed for the remaining period of observed values, and goodness of fit between recorded and simulated values is reassessed. This type of split-data set calibration/validation procedure will be followed for the Housatonic River modeling.

The final check of the modeling results will be performed using a "weight of evidence" approach, evaluating the results obtained from the model together with an objective analysis of the data and other related studies, including those summarized in Section 6. This is a key component of the overall conceptual modeling approach described in Section 3.

5.2 HSPF CALIBRATION AND VALIDATION PROCEDURES

The application of HSPF to the Housatonic River watershed will follow the standard model application procedures as described in the HSPF Application Guide (Donigian et al., 1984), in numerous watershed studies over the past 15 years (see Bibliography for HSPF [Donigian, 1999]), and recently in the HSPF application to the Chesapeake Bay Watershed (Donigian et al., 1994). Model application procedures for HSPF include database development, watershed segmentation, (discussed in the previous sections) and hydrology, sediment, and water quality calibration and validation.

5.2.1 Model Calibration

Model calibration is required for parameters that cannot be deterministically evaluated from topographic, climatic, edaphic, or physical/chemical characteristics. The majority of HSPF parameters are not in this category. Calibration will be based on several years of simulation (at least 3 to 5 years) in order to evaluate parameters under a variety of climatic, soil moisture, and water quality conditions. The areal variability of meteorologic data series, especially precipitation and air temperature, may introduce some as yet unknown level of uncertainty in the simulation; this will be evaluated explicitly during the calibration process. Years with heavy precipitation are often better simulated because of the relative uniformity of large events over a watershed. In contrast, low annual runoff may be caused by a single or a series of small events that did not have a uniform areal coverage. Parameters calibrated on a dry period of record may not adequately represent the processes occurring during the wet periods. Also, the effects of initial conditions of soil moisture and pollutant accumulation can extend for several months beyond the period of record in which the data were collected, resulting in biased parameter values calibrated on short simulation periods. Calibration should result in parameter values that produce the best overall agreement between simulated and observed values throughout the calibration period.

When modeling land surface processes, hydrologic calibration (runoff and streamflow) must precede sediment and water quality calibration because runoff is the transport mechanism by which nonpoint loadings occur. Likewise, adjustments to the instream hydraulics simulation must be completed before instream sediment and water quality transport and processes are calibrated.

5.2.2 Model Validation

Model performance and validation will be evaluated through graphical, quantitative, qualitative, and statistical measures.

For flow simulations where continuous records are available, all these techniques will be used during both the calibration and validation phases. Comparisons of simulated and observed values will be performed for daily, monthly, and annual values, in addition to flow-frequency duration assessments. Statistical procedures will include correlation and model-fit efficiency coefficients.

For sediment and water quality constituents, model performance will be based primarily on visual and graphical presentations because the frequency of observed data is often inadequate for accurate statistical measures. However, alternative model performance assessment techniques consistent with the population of observed data available for model testing are discussed and described in the Modeling Study QAPP (Beach et al., 2000).

5.3 EFDC CALIBRATION AND VALIDATION PROCEDURES

EFDC will be used to model the Housatonic River's hydrodynamics, sediment transport, and abiotic PCB transport. The calibration/validation of EFDC will be dependent on the final HSPF model. EFDC will obtain boundary conditions from HSPF for water quantity and timing (i.e., flow rates), sediment loads, and tributary PCB loadings.

It is important to conduct the modeling in sequential steps for both the calibration and validation periods. A logical and sequential process must be followed that first resolves the hydrodynamics, then the sediment transport, and finally the abiotic PCB parameters. The steps of the calibration/validation process, and the parameters upon which the focus will be placed at each step are shown in Table 5-1. A fundamental principle underlying the calibration/validation process is the need to maintain mass balance for every constituent. This will be evaluated early in each step of the process to ensure that there are no unaccounted-for gains or losses.

Additional calibration/validation efforts will be necessary for the Woods Pond submodel as this may include implementation of a 3-D component of EFDC to capture the Woods Pond dynamics. The additional focus will be on the vertical profiles of temperature, dissolved oxygen, and pH.

Table 5-1

EFDC Model Calibration/Validation Steps

Step	Parameters of Focus
Initial Condition Development	Bathymetry, sediment distribution, PCB distribution, roughness conditions
Inflow Development	HSPF-linked results to each upstream condition, point tributary, and local runoff/distributed inflows/loadings
Hydraulic Head/Stage Comparisons	Comparison of New Lenox and Woods Pond elevations and timing
Velocity Comparisons	Comparisons to manual velocity measurements and ADCP velocities
Sediment Transport	Comparisons of total loads, areal distribution of depositional/erosional areas, and Woods Pond deposition rates
Abiotic PCBs	Comparisons of water column fluxes/concentrations, distribution of high/low-concentration areas

5.3.1 Model Calibration

EFDC will be calibrated for hydrodynamics, sediment transport, and abiotic PCBs to the periods that have the highest density and quality of data. The storm event data collected during 1999 meet these criteria and will be a fundamental basis for calibration. Ten storms were monitored with seven of these events having the full protocol achieved for sampling and analysis. This data set provides a good basis for developing short-term event calibrations. The steps outlined in Table 5-1 will be followed for each storm event. However, the initial conditions may not be changed from one storm event to the next, depending on a detailed evaluation of the data. The final step of the calibration process will be a 2-year simulation from early 1999 to the present to evaluate the intermediate time scale processes. An iterative approach may be necessary to apply new parameterizations to prior periods to ensure consistency and achieve calibration tolerances defined in the Modeling Study QAPP.

The calibration will include comparisons of averaged (time and space) model output to measured data. The measured data will be preprocessed to similar time and space scales that are appropriate for the model scale. Depending on the nature of the event/period, model results may be averaged into hourly, daily, yearly, or decadal time scales. Spatial scales may be points, increments of river miles, or spatially averaged segments, depending on the nature of the parameter. For example, the evaluation of the total sediment load passing New Lenox Road will vary in scale from that used for determining wetland PCB distributions. Specific metrics that will be used in calibration are provided in the Modeling Study QAPP (Beach et al., 2000).

5.3.2 Model Validation

The validation starting point will be 1979-1980. Data collected by USGS and the Connecticut Agricultural Experimental Station in 1979 and data collected by Stewart (Stewart Laboratories, 1982) during the 1980-1982 timeframe are being evaluated for usability. It is expected that one or the other will be used to set the initial conditions of the model. The validation will include a period of approximately 20 years, ending in 2000. The same process will be followed as that used in calibrations, with only the date of initial conditions and duration of the simulation being different. It is possible that the longer term simulation performed during the validation will detect some divergence from the observed data and will require some model adjustments. Any revision to a "calibrated" model during the validation step will necessarily require recalibration. However, the parameters (or magnitude of change) that would be modified based upon the results obtained during validation are not likely to have a noticeable effect on the calibration results representing a shorter time period. Further discussion of the comparisons that will be made to evaluate the validation effort are provided in the Modeling Study QAPP (Beach et al., 2000). Ultimately, the model is expected to reasonably represent the entire range of observed data, at which point it will be considered validated and ready for predictive simulations and alternative analyses.

5.4 AQUATOX CALIBRATION AND VALIDATION PROCEDURES

The philosophy of the application of AQUATOX, emphasizing generality and reality, is one that has been used for the past 25 years by Park and colleagues (Park et al. 1974; Collins, 1980; Park et al. 1981; Park and Collins, 1982). Because AQUATOX is to be applied to changing conditions at the ecosystem level in the Housatonic River, it must be general, and it should be realistic in its representation of both the ecosystem and the fate of PCBs in that system. Taken together, these characteristics represent a measure of accuracy. The goal will be the simulation of biotic and contaminant behavior with a robust set of parameters, some that are site-specific and some that are independent of site conditions. While precise matches between model predictions and observed data are not expected, comparisons will be made to ensure that the predictive capability of the model produces results that are reasonable when evaluated against site-specific data. This evaluation will be performed within the framework of ecological variability through space and time, particularly at the scale needed to address the purposes of the model in evaluating the response associated with remedial alternatives when compared with baseline conditions.

There is a long history of development and testing of the aquatic ecosystem formulations that are embodied in AQUATOX, and application to diverse aquatic systems continues. Literature on the fate of PCBs on the toxicity of specific congeners, site-specific data, and independent data sets derived from other contaminated sites (see Appendix D) provide an excellent basis for parameterizing and testing the process-level chemodynamic formulations in AQUATOX. Calibration will involve iterative parameterization and testing of river ecosystem and PCB fate and bioaccumulation constructs. Validation or verification of the model implementation for the Housatonic River will involve comparison with existing site data to ensure that the model results represent the known trends in the fate and effects of PCBs. The model will then provide the capability to forecast future behavior of PCBs in the Housatonic River, given changing conditions under various remediation alternatives.

5.4.1 River Ecosystem Calibration

The PSA includes shallow and deep reaches and backwater areas of the Housatonic River, and shallow and deep segments of Woods Pond. Proposed biotic state variables representing the complex food web include periphyton, phytoplankton, macrophytes, filamentous algae/duckweed, invertebrates (cladocerans, mayflies, caddisflies, dragonflies, midges, worms, and crayfish), and fish (minnows, goldfish and carp, brown bullhead, white suckers, sunfish, yellow perch, and largemouth bass).

Because of extensive previous applications to impoundments, AQUATOX will represent the Woods Pond ecosystem with little additional calibration necessary. Application to the river ecosystem also will be relatively straightforward. AQUATOX has a large database of parameters representing many riverine invertebrates and fish; therefore, only minor adjustments for calibration are anticipated. In addition, the generality of the model in representing periphyton, macrophytes, various invertebrates, and fish in the river will be tested using available data from other streams and small rivers, in addition to data from the Housatonic. In particular, published data from East Poplar Creek and Walker Branch, Tennessee, and the Little Miami River, Ohio, will be used to further augment the river implementation. The goal is to represent the ecosystem and food web of the Housatonic River realistically so that dietary exposure and bioenergetics of the invertebrates and fish can be used to predict fate and bioaccumulation of PCBs. Biomagnification of hydrophobic compounds such as PCBs is sensitive to the number of trophic levels and to the structure of detritus-based and plant-based food webs, so it is important to represent the complexity of the Housatonic biota, given the available data and general principles of aquatic ecology.

5.4.2 PCB Calibration

The goal is to model PCB homologs and three or more selected congeners in sufficient detail so that the selective microbial degradation and volatilization of homologs and congeners can be predicted, as well as the selective bioaccumulation and biotransformation. The first step is to parameterize and, as necessary, modify fate and effects formulations to best represent PCBs in the Housatonic River. Process-level equations will be tested against experimental data available in the literature (see Appendix D). Simulations will be run using newly collected PCB data, particularly congener data, from the Housatonic River. Similarly, published (Hill and Napolitano, 1997) and unpublished congener data from East Fork Poplar Creek, Tennessee (which is similar to the Housatonic River in several respects) will be used to further refine the model. Sensitivity analyses will be run to determine which parameters have the most effect on the simulations. If the model is inappropriately sensitive to a parameter, then the formulation will be reconsidered and modified if necessary. Sensitive parameters will be noted for use in uncertainty analyses in later simulations.

The second phase of calibration will involve running the distributed version of AQUATOX in tandem with the EFDC and HSPF models to test and modify the hydrodynamic and sediment linkages and their applicability to modeling PCB transport, sedimentation, burial, bioturbation, and resuspension.

5.4.3 AQUATOX Validation

Validation will be performed using Housatonic data starting at the beginning of the period of record until present; only a subset of these data will be used for calibration. Observed data for the earlier years include total PCBs and Aroclors in sediments and fish. Given the limited historical PCB data, particularly for biota, and absence of historical congener data at the site, the validation process will be based on congener distributions in Aroclors 1254 and 1260 and the congener patterns observed in the site-specific data. An effort is underway to evaluate the analytical methodologies used in developing the various data sets during the period of record to determine how best to adjust the data, if needed or possible, to provide comparability between data sets.

Because of long half-lives of the more chlorinated PCBs in adult fish and slow degradation rates in sediments, a long simulation period (1979 to 2000) will be used for validation. This period includes the calibration period, but, with initiation 15 years earlier, will provide an independent test with the high-quality data available from more recent studies.

There are several measures of model performance that can be used (Bartell et al., 1992; Schnoor, 1996). The difficulty is in comparing general model behavior over long periods—with rapid fluctuations due to natural occurrences such as storm events and algal blooms, seasonal fluctuations, and annual variability—to observed data from a few points in time with poorly defined sample variability. Recognizing that the evaluation process is limited by the quantity and quality of data, stringent measures of goodness of fit are inappropriate; therefore, a sequence of tests will be used to evaluate the calibration and validation of the model. Further details are provided in the Modeling Study QAPP (Beach et al., 2000).

5.5 PROPOSED CALIBRATION AND VALIDATION PERIODS

Based on the review to date, the recommendations for calibration and validation periods are presented in Table 5-2.

Table 5-2

Calibration and Validation Periods

	Calibration	Validation
Streamflow	1991-2000	1979-2000
Water temperature	1991-2000	1979-2000
Sediment loads	1991-2000	1979-2000
Nonpoint loads (nutrients/BOD/organics)	1996-2000	1979-2000
Stage height	1999-2000	1979-2000
Velocity	1999-2000	suitable data not available
Suspended solids (water column)	1999-2000	1979-2000
Sediment bed solids	1999-2000	1979-2000
PCBs (water column/bed)	1999-2000	1979-2000
PCBs (fate and bioaccumulation)	1995-2000	1979-2000

Note: The validation period uses the longest period of time and is bounded by available data. This approach allows use of the longest timeframe for which model performance can be evaluated. The resulting validated model is more suitable for evaluating the model's predictive capability for simulating baseline conditions and the long-term effects of potential remedial alternatives.

Two basic types of data will be used for model calibration. The first type is continuously recorded data (i.e., flow records) and the second type is periodic/episodic data (i.e., all other data). The HSPF model uses both the first and second types of data, but particularly relies upon the continuous streamflow records for the hydrology calibration. The EFDC and the AQUATOX models primarily use the second type of data for calibration/validation.

For the hydrologic data records, the most important factor in selecting a calibration and validation period (assuming the data are available and representative) is the need to represent a range of hydrologic conditions. Calibrating to a period of record that includes extremes such as large storm events and long droughts would be ideal. The goal would then be to validate to another period that has similar variability. The dates selected for the calibration/validation listed above are expected to achieve this goal.

Due to the relatively limited availability of the episodic data throughout the long time period required for this modeling effort (more than 50 years), it is preferable to use as much of the available data as possible during validation. The use of the long-term period for validation will allow the evaluation of trends expected over longer time periods in the modeling results that cannot be adequately assessed using the shorter time periods in the calibration process. In order to achieve the longest validation period for this modeling study, the validation period will extend to present-day, incorporating the calibration period.

The dates listed in Table 5-2 represent data from reports prepared by numerous authors and produced for various purposes. The 1998-2000 (WESTON, 2000a) data set is the only one that was specifically collected for the purposes of the modeling presented in this document. Not every data set will be useful for each modeling effort. For example, PCB fate calibration of AQUATOX will use the 1995 Smith and Coles data (Smith and Coles, 1997), among others, but these data (tissue residue concentrations) will not be useful for EFDC and HSPF. The Modeling Study QAPP (Beach et al., 2000) provides a complete summary of the general application of each data set.

5.6 SENSITIVITY/UNCERTAINTY ANALYSES

5.6.1 Sensitivity/Uncertainty Analyses for EFDC and HSPF

The computational demands of both EFDC and HSPF are such that formal probabilistic analyses with numerous iterations of simulation runs are not readily feasible. Consequently, the approach for these system components will be to perform sensitivity analyses on selected model parameters and boundary conditions that are known to be critical based on past experience with both models. The focus will be twofold: (1) evaluate the impacts of key calibration parameters on both process representations and critical flux input from EFDC and HSPF to AQUATOX, and (2) develop a basis for selecting an appropriate distribution of loadings and fluxes from EFDC/HSPF for use in AQUATOX uncertainty analyses.

For HSPF, the analyses will focus on the representation of the transport and nonpoint load generation processes in the model; evaluating the sensitivity of the PCB loads to variations in the critical model parameters. These transport parameters will include those related to both runoff generation (i.e., infiltration, soil moisture capacity, surface characteristics) and sediment erosion (i.e., soil erodibility, vegetative cover). It is expected that, due to the nature of historical loading of PCBs at this site, the source characterization will be one of the most uncertain aspects of the PCB loading simulations, and therefore a focus of the sensitivity analyses. Alternative approaches will be investigated for quantifying the PCB loads, ranging from the use of simulated flows and observed PCB concentrations to direct modeling and calibration of PCBs as a function of surface runoff and sediment loads. Each alternative approach will be subjected to sensitivity analyses as part of the assessment.

For EFDC, the focus will be on the hydrodynamic and sediment transport parameters that control the corresponding processes. The primary hydrodynamic parameters of concern are the bottom roughness and the vegetation resistance to flow, and the spatial variation of these throughout the model domain. The sediment transport parameters that will be included in the sensitivity analysis include, but are not necessarily limited to, critical shear stress governing resuspension and deposition, terms relating time-varying dry bed density to bed shear strength (during cohesive bed consolidation), and effective settling velocities for both noncohesive and cohesive sediment.

In addition to the evaluation of sensitivity within the component models, the sensitivity of the model linkages will be evaluated in terms of the data transfers among the models. This will include the associated assumptions of time and spatial aggregation involved in processing output of one model for input to the other, and the impacts of these assumptions (as discussed previously in Section 4.4) on the model predictions.

5.6.2 Sensitivity/Uncertainty Analyses for AQUATOX

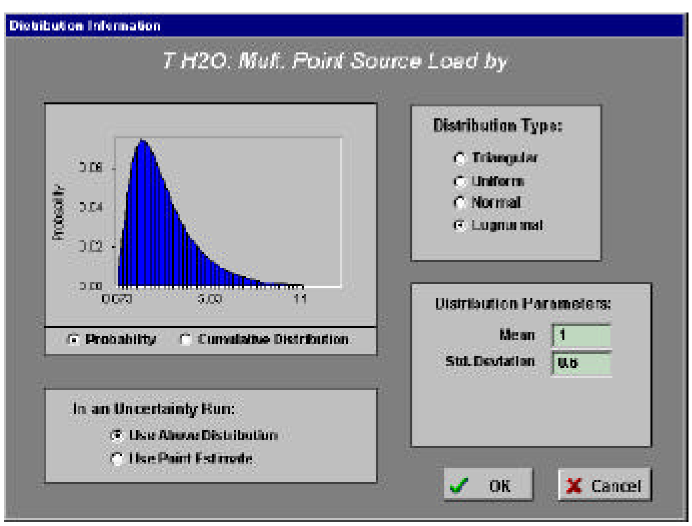
Sensitivity analyses will be performed on critical biotic parameters such as maximum consumption and natural mortality rates, which vary among experiments and representative organisms. Also of interest are the relative effects of physicochemical characteristics such as octanol-water partition coefficients and Henry’s Law constants that cannot be measured easily;

and chemodynamic parameters such as microbial degradation, which may be subject to both measurement errors and poorly determined environmental controls. These will be evaluated during the calibration phase and they may influence the formulation and parameterization of the model.

The effects of uncertain inputs and natural variability on the model predictions are important and will be evaluated. Uncertainty analysis also will consider sources of uncertainty and variation inherent in natural systems and with regard to contaminants (specifically PCBs). These include: site characteristics such as water depth; environmental loadings such as water flow, temperature, and light, which may have a stochastic (random) component; and the characterization of pollutant loadings from runoff and point sources, which may vary stochastically from day to day. The aggregate effect of these components on the simulation results will be examined.

Probabilistic modeling approaches may be used as tools for evaluating the implications of uncertainty in the analyses. In this modeling study, AQUATOX provides this capability by allowing the user to specify the types of distribution and key statistics for a wide selection of input variables. Depending on the specific variable and the amount of available information, any one of several distributions may be the most appropriate. A lognormal distribution is the default for environmental and contaminant loadings; distributions for constant loadings can be sampled daily, providing day-to-day variation; distributions for dynamic loadings, which will drive the AQUATOX simulations, use multiplicative factors that can be sampled once each simulation (Figure 5-1).

Figure 5-1 Distribution Screen for Point-Source Loading of Toxicant in Water



A sequence of increasingly defined distributions should be considered for most parameters. If only two values are known and nothing more can be assumed, the two values may be used as minimum and maximum values for a uniform distribution (Figure 5-2). If minimal information is available but there is reason to accept a particular value as most likely, perhaps based on calibration, then a triangular distribution may be most suitable (Figure 5-3); note that the minimum and maximum values for the distribution are constraints that have zero probability of occurrence. If additional data are available indicating both a central tendency and spread of response, then a normal distribution (Figure 5-4) may be most appropriate. All distributions are truncated at zero because negative values would have no meaning.

Figure 5-2 Uniform Distribution for Henry's Law Constant for Esfenvalerate

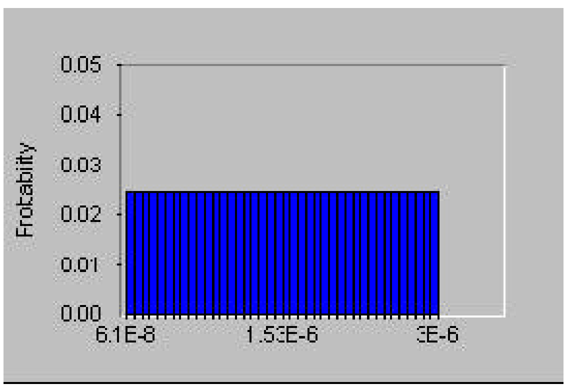
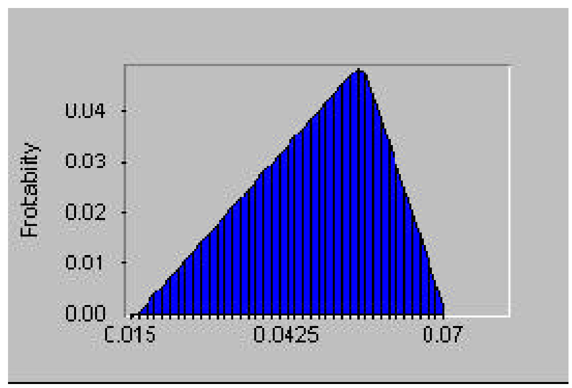


Figure 5-3 Triangular Distribution for Maximum Consumption Rate for Bass



Efficient sampling from the distributions is obtained with the Latin hypercube method (McKay et al., 1979; Palisade Corporation, 1991). This procedure is used by AQUATOX with algorithms originally written in FORTRAN (Anonymous, 1988). Depending on how many iterations are chosen for the analysis, each cumulative distribution is subdivided into that many equal segments. Then a uniform random value is chosen *within* each segment and used in one of the subsequent simulation runs (Figure 5-5). This method is particularly advantageous because all regions of the distribution, including the tails, are sampled. The default is 20 iterations, meaning that 20 simulations will be performed with sampled input values; this should be considered the minimum number to provide any reliability. The optimal number can be determined experimentally by noting the number required to obtain convergence of mean response values for key state variables; i.e., at what point do additional iterations not result in significant changes in the results? As many variables may be represented by distributions as desired, but the method assumes that they are independently distributed. By varying one parameter at a time the sensitivity of the model to individual parameters can be determined.

Figure 5-4 Normal Distribution for Maximum Photosynthetic Rate for Diatoms

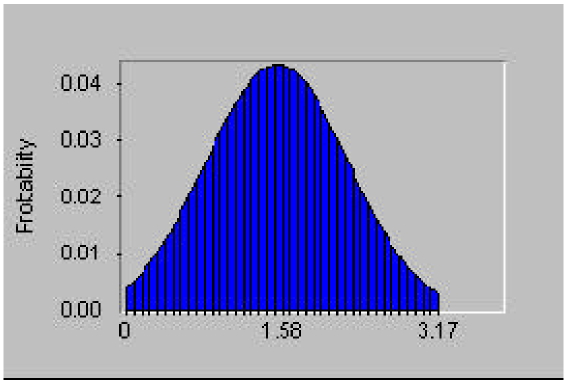


Figure 5-5 Latin Hypercube Sampling of a Cumulative Distribution with a Mean of 25 and Standard Deviation of 8 Divided into 5 Intervals